Optimized Neural Networks to Search for Higgs Boson Production at the Tevatron.

E. Boos, L. Dudko [a]

[a]Scobeltsyn Institute of Nuclear Physics of Moscow State University,
Vorobyevy Gory, 119899, Moscow, Russia

An optimal choice of proper kinematical variables is one of the main steps in using neural networks (NN) in high energy physics. Our method of the variable selection is based on the analysis of a structure of Feynman diagrams (singularities and spin correlations) contributing to the signal and background processes. An application of this method to the Higgs boson search at the Tevatron leads to an improvement in the NN efficiency by a factor of 1.5-2 in comparison to previous NN studies.

1. The basic idea

In High Energy physics a discrimination between a signal and its corresponding backgrounds by Neural Networks (NN) is especially remarkable when the data statistics are limited. In this case it is important to optimize all steps of the analysis. One of the main questions which arises in the use of NNs is which, and how many variables should be chosen for network training in order to extract a signal from the backgrounds in an optimal way. The general problem is rather complicated and finding a solution depends on having a concrete process for making the choice, because usually it takes a lot of time to compare results from different sets of variables.

One observation which helps in making the best choice of the most sensitive variables is to study the singularities in Feynman diagrams of the processes. Let us call those kinematic variables in which singularities occur as "singular variables". What is important to stress here is that most of the rates for both the signal and for the backgrounds come from the integration over the phase space region close to these singularities. One can compare the lists of singular variables and the positions of the corresponding singularities in Feynman diagrams for the signal process and for the backgrounds. It is obvious that if some of the singular variables are different or the positions of the singularities are different for the same variable for the signal and for the backgrounds the corresponding distributions will differ most strongly. Therefore, if one uses all such singular variables in the analysis, then the largest part of the phase space where the signal and backgrounds differ most will be taken into account. One might think that it is not a simple task to list all the singular variables when the phase space is very complex, for instance, for reactions with many particles involved. However, in general, all singular variables can be of only two types, either s-channel: $M^2_{f1,f2} = (p_{f1} + p_{f2})^2$, where p_{f1} and p_{f2} are the four momenta of the final particles $f1$ and $f2$ or t-channel: $\hat{t}_{i,f} = (p_f - p_i)^2$, where p_f and p_i are the momenta of the final particle (or cluster) and the initial parton. For the $\hat{t}_{i,f}$ all the needed variables can be easily found in massless case: $\hat{t}_{i,f} = -\sqrt{\hat{s}}e^Y p_T^f e^{-|y_f|}$, where $\hat{s}$ is the total invariant mass of the produced system, and Y is the rapidity of the total system (rapidity of the center mass of the colliding partons), p_T^f and y_f are transverse momenta and pseudorapidity of the final particle f. The idea of using singular variables as the most discriminative ones is described in [1] and the corresponding method was demonstrated in practice in [2].

Singular variables correspond to the structure of the denominators of Feynman diagrams. Another type of interesting variables corresponds to the numerators of Feynman diagrams and reflects the spin effects and the corresponding difference in angular distributions of the final particles. In order to discriminate between a signal and the backgrounds, one should choose in addition to singular variables mentioned above those angular variables whose distributions are different for the signal and backgrounds. The set of these singular and angular variables will be the most efficient set for a NN analysis.

The third type of useful variables which we called "Threshold" variables are related to the fact that various signal and background processes may have very different thresholds. Therefore the distributions over such kind of variables also could be very different keeping in mind that effective parton luminosities depend strongly on $\hat{s}$. The variable $\hat{s}$ would be a very efficient variable of that kind. However, the problem is that in case of neutrinos in the final state one can not measure $\hat{s}$ and should use the effective $\hat{s}$ which is reconstructed by solving t-,W-mass equations for the neutrino longitudinal momenta. That is why we propose to use not only the effective variable $\hat{s}$ but the variable H_T^{jets} as well.

To apply the method it is important to use a proper Monte-Carlo model of signal and background events which includes all needed spin correlations between production and decays. For the following analysis we have calculated the complete tree level matrix elements for the background processes with all decays and correlations by means of the CompHEP program [3]. The corresponding events are available at the FNAL Monte-Carlo events database [4].

2. Applying the method

The present estimation of the expected sensitivities for the light Higgs boson search at the Tevatron by means of NNs is given in [5]. Based on the method described above we improve the efficiency of the NN technique. In the analysis we choose the Higgs boson mass to be $M_H = 115$ GeV. We model the detector smearing by the SHW package [6].

First of all we exclude ineffective variables from the old set [5], like P_T^e from the W-boson (shown at the left plot in Fig. 1). After the corresponding analysis of Feynman diagrams and comparison of kinematical distributions we added the new variables for NN training. The example distribution for the new variable $(cos(z_{axis}, e))$ is shown in the right plot of Fig. 1. At the next

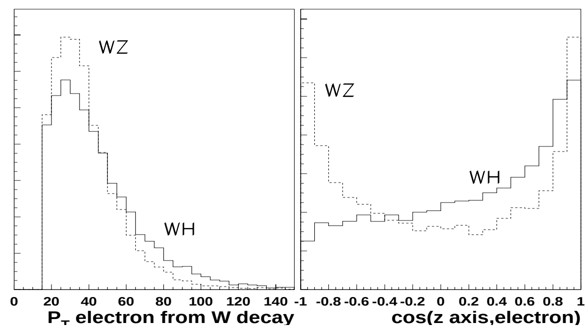

Figure 1. Examples of the old kinematic variable (left plot) and the new one (right plot)

step we constructed the set of NNs for pairs of the signal (WH) and each of the background from the complete set of principle backgrounds ($Wb\bar{b}$, WZ, $t\bar{t}$, $tb(j)$).

The standard steps of NN training were used for the NNs with the old set of input variables and with the new one. Efficiencies of networks with different sets have been compared based on the criteria that for the better net the "Error function" $E = \frac{1}{N}\sum_{i=1}^{N}(d_i - o_i)^2$, where d_i and o_i are the desired and real outputs of the net and N is the number of test events, is smaller. Two examples of distributions you can see in Fig 2 for the $WH - t\bar{t}$ network (left plot) and $WH - WZ$ network (right plot). One can see a significant improvement for the networks with new input sets in comparison with old sets of variables, since the corresponding curves of the error function are significantly lower.

3. Results

Based on the described method we have constructed the new NNs to search for a light Higgs boson at the Tevatron. After checking the improvement in efficiency of new networks we recommend the new sets of input variables for NNs, which are shown below:

- $Wb\bar{b} - WH$
  NN: $M_{b\bar{b}},\ P_T^{b1},\ P_T^{b2},\ P_T^{bb},\ \hat{s},\ H_T^{jets},$
  $cos(b1,b2)|_{lab},\ cos(b1,b1b2)|_{b1b2}$

- $WZ - WH$
  NN: $M_{b\bar{b}},\ P_T^{b1},\ P_T^{b2},\ H_T^{jets},$
  $cos(b1,b2)|_{lab},\ Q\times cos(z,b1)|_{lab},\ cos(W,e)|_W$

- $t\bar{t} - WH$
  NN: $M_{b\bar{b}},\ M_{Wb},\ \hat{s},\ M_{Wjets-b},\ H_T^{jets},$
  $Q\times cos(\psi_{axis},e)|_{top},\ cos(b1,b1b2)|_{b1b2}$

- $tbj,\ tb - WH$
  NN: $M_{b\bar{b}},\ M_{Wb},\ P_T^{b2},\ \hat{s},$
  $M_{Wjets-b},\ P_T^{top},\ H_T^{jets},\ cos(z,e)|_{lab},$
  $Q\times cos(z,b1)|_{top},\ cos(e,j)|_{top}$

where there are three types of variables:

- "Singular" variables (denominator of Feynman diagrams):
  M_{12} is the invariant mass of two particles and/or jets (1 and 2) and corresponds to s-channel singularities;
  P_T^f (the transverse momenta of *f*);
  $M_{Wjets-b}$ is the invariant mass of the W and all jets except the b-jet for which the $M_t = (p_W + p_b)$ is closest to the top quark mass;

- "Angular" variables (numerator of Feynman diagrams, spin effects):
  $cos(b1,b1b2)|_{b1b2}$ means the cosine of the angle between highest P_T b-quark and vector sum of the two highest P_T b-quarks in the rest frame of these two b-quarks. Scalar (Higgs) and vector (gluon, Z-boson) particle decays lead to significantly different distributions on this variable, this is also very much different for the case when b-quarks come from the decay of top and anti-top quarks;
  $cos(b1,b2)|_{lab}$ characterizes how much two b-quarks are collinear;
  $cos(z,b1)|_{lab}$ and $cos(W,e)|_W$ reflect the difference in t-channel Z-boson and s-channel Higgs-boson production topologies where $_{lab}$ means the laboratory rest frame and z means the z-axis;
  $cos(\psi_{axis},e)|_{top}$ [7] and $cos(e,j)|_{top}$ [8] are the top quark spin correlation variables used in the analysis of the top quark pair and single production, the lepton charge Q is added here to take uniformly into account the electron and the positron contributions from the W-boson decays.

- "Threshold" variables. As explained above the $\hat{s}$ and H_T^{jets} variables are used in our analysis.

As one can see from the Fig.2 using the new NN variables allows to improve the NN efficiency by a factor of 1.5-2 depending on the background process. It will lead to corresponding improvement in prospects to find a light Higgs at the Tevatron. However, one needs to take into account the ZH production channel as well as a number of detector efficiencies in order to predict a realistic discovery limit.

Acknowledgements

The work was partly supported by the INTAS 00-0679, CERN-INTAS 99-377, YSF 02/239 and Universities of Russia UR.02.03.002 grants. E. Boos thanks the Humboldt Foundation for the Bessel Research Award.

REFERENCES

1. E. Boos, L. Dudko and T. Ohl, Eur. Phys. J. C **11**, 473 (1999); E.E. Boos and L.V. Dudko, "Singularities of Feynman Diagrams and Optimal Kinematic Variables for Neural Networks", DØ Note 3612, February 1999.
2. L. Dudko, In Proceedings of AIHENP'99, Crete, April, 1999; V. M. Abazov *et al.* [D0

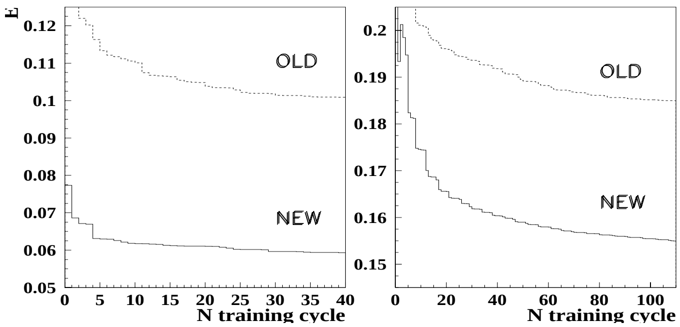

Figure 2. NN Error function for the $WH-t\bar{t}$ (left plot) and $WH-WZ$ networks (right plot).

Collaboration], "Search for single top quark production at D0 using neural networks," Phys. Lett. B **517**, 282 (2001)

3. A. Pukhov *et al.*, hep-ph/9908288; E.E. Boos *et al.*, INP MSU 94-36/358 and SNUTP-94-116, hep-ph/9503280;
4. `http://www-d0.fnal.gov/~dudko/mcdb/`
5. P.C. Bhat, R. Gilmartin, H.B. Prosper Phys.Rev. D62 (2000) 074022; hep-ph/0006162
6. `http://www.physics.rutgers.edu/~jconway/ /soft/shw/shw.html`
7. Gregory Mahlon, McGill/98-34, hep-ph/9811281
8. G. Mahlon and S. Parke, Phys. Rev. **D55** (1997) 7249.
   G. Mahlon and S. Parke, Phys. Lett. **B476**, (2000) 323.
   E.E. Boos, A.V. Sherstnev Phys. Lett. **B534**, (2002) 97-105